On Minimizing the MSE in the Presence of Channel State Information Errors

Gábor Fodor*†, Piergiuseppe Di Marco*, Miklós Telek‡
*Ericsson Research, Stockholm, Sweden. E-mail: gabor.fodor@ericsson.com
†Royal Institute of Technology, Stockholm, Sweden. E-mail: pidm@kth.se
‡Budapest University of Technology and Economics, Budapest, Hungary. E-mail: telek@hit.bme.hu

***Abstract*—We consider the uplink of a multiuser multiple input multiple output (MU MIMO) system, in which the base station acquires channel state information (CSI) for which the estimation error depends on the resources assigned to the uplink pilot symbols. For this system, we first derive the receiver that minimizes the mean square error (MSE) of the uplink detected data symbols, as opposed to the *naïve* receiver that does not minimize the MSE in the presence of CSI errors. We then derive a closed form expression for the MSE as a function of the employed pilot-to-data power ratio, number of antennas and the MU MIMO interference power. This expression allows us to gain the insight that the gain of using the actual MMSE receiver as opposed to the naïve receiver becomes particularly important when the number of BS antennas is large.**

I. Introduction

In multiuser multiple input multiple output (MU MIMO) systems, the fundamental trade-off between spending resources on channel state information (CSI) acquisition and data transmission is known to affect the performance in terms of spectral and energy efficiency [1], [2]. Therefore, balancing the pilot-to-data power ratio (PDPR) [3] and determining the number of pilot and data symbols are important aspects of designing MIMO systems [4], [5], [6]. From a different perspective, a related work combined a transmitter employing a linear dispersion code (LDC) and a linear minimum mean square error (MMSE) detector at the receiver [7]. It has been found that optimizing the average normalized mean square error (MSE) is relevant for detectors employing a linear front end and helps designing optimal transmit strategies. In this paper we consider the uplink of a MIMO system employing an MMSE receiver for data reception [8]. The MMSE receiver is initialized by the estimates of the channel state information rather than assuming the availability of perfect CSI. Thus, our contribution to the existing literature is two-fold:

1) We derive the actual MMSE receiver that, – in contrast to the classical or *naïve* formula [9] – minimizes the MSE of the estimated uplink data symbols in the presence of PDPR dependent estimation errors.
2) Secondly, we derive a closed form exact expression for the MSE, as a function of not only the PDPR but also the number of antennas. This exact formula allows us to arrive at the key insight that employing the actual MMSE gives large gains as the number of antennas grows large.

II. System and Channel Estimation Model

We consider the uplink of a MU MIMO system, in which the mobile stations (MS) transmit orthogonal pilot sequences $\mathbf{s} = \left[s_1, ..., s_{\tau_p}\right]^T \in \mathbb{C}^{\tau_p \times 1}$, in which each pilot symbol is scaled as $|s_i|^2 = 1$, for $i = 1, .., \tau_p$. The pilot sequences are constructed such that they remain orthogonal as long as the number of spatially multiplexed users is maximum τ_p. Specifically, without loss of generality, we assume that the number MU-MIMO users is $K \leq \tau_p$. In practice, $K \ll N_r$, where N_r is the number of antennas at the BS.

In this paper we assume a comb type arrangement of the pilot symbols [10]. Given F subcarriers in the coherence bandwidth, a fraction of τ_p subcarriers are allocated to the pilot and $F_d = F - \tau_p$ subcarriers are allocated to the data symbols. Each MS transmits at a constant power P_{tot}, however, the transmission power can be distributed unequally in each subcarrier. In particular, considering a transmitted power P_p for each pilot symbol and P for each data symbol transmission, the sum constraint of $\tau_p P_p + (F - \tau_p)P = P_{tot}$ is enforced. Thus, the $N_r \times \tau_p$ matrix of the received pilot signal from a specific MS at the BS can be conveniently written as:

$$\mathbf{Y}^p = \alpha\sqrt{P_p}\mathbf{h}\mathbf{s}^T + \mathbf{N}, \tag{1}$$

where we assume that $\mathbf{h} \in \mathbb{C}^{N_r \times 1}$ is a circular symmetric complex normal distributed column vector with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}$ (of size N_r), denoted as $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, α accounts for the propagation loss, $\mathbf{N} \in \mathbb{C}^{N_r \times \tau_p}$ is the spatially and temporally additive white Gaussian noise (AWGN) with element-wise variance σ_p^2, where the index p refers to the noise power on the received *pilot* signal.

In this paper we assume that the BS uses the popular least square (LS) estimator that relies on correlating the received signal with the known pilot sequence. Note that our methodology to determine the MSE of the received data is not confined to the LS estimator, but is directly applicable to an MMSE or other channel estimation techniques as well. For each MS, the BS utilizes pilot sequence orthogonality and estimates the channel based on (1) assuming:

$$\hat{\mathbf{h}} = \mathbf{h} + \mathbf{w} = \frac{1}{\alpha\sqrt{P_p}}\mathbf{Y}^p\mathbf{s}^*(\mathbf{s}^T\mathbf{s}^*)^{-1} = \mathbf{h} + \frac{1}{\alpha\sqrt{P_p}\tau_p}\mathbf{N}\,\mathbf{s}^*, \tag{2}$$

where $\mathbf{s}^* = \left[s_1^*, ..., s_{\tau_p}^*\right]^T \in \mathbb{C}^{\tau_p \times 1}$ denotes the vector of pilot symbols and $(\mathbf{s}^T\mathbf{s}^*) = \tau_p$. By considering $\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, it follows that the estimated channel $\hat{\mathbf{h}}$ is a circular symmetric complex normal distributed vector $\hat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$, with

$$\mathbf{R} \triangleq \mathbb{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\} = \mathbf{C} + \frac{\sigma_p^2}{\alpha^2 P_p \tau_p}\mathbf{I}_{N_r}. \tag{3}$$

As it was shown in [10], the distribution of the channel realization $\mathbf{h}$ conditioned on the estimate $\hat{\mathbf{h}}$ is normally distributed as follows:

$$(\mathbf{h} \mid \hat{\mathbf{h}}) \sim \mathbf{D}\hat{\mathbf{h}} + \mathcal{CN}\Big(\mathbf{0}, \mathbf{Q}\Big), \tag{4}$$

where $\mathbf{D} \triangleq \mathbf{C}\mathbf{R}^{-1}$ and $\mathbf{Q} \triangleq \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}$.

III. Linear MMSE Receiver

A. Received Data Signal Model

The MU-MIMO received data signal at the BS can be written as:

$$\mathbf{y} = \underbrace{\alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} x_\kappa}_{\text{User-}\kappa} + \underbrace{\sum_{k \neq \kappa}^{K} \alpha_k \mathbf{h}_k \sqrt{P_k} x_k}_{\text{Other users}} + \mathbf{n}_d, \tag{5}$$

where $\alpha_k \cdot \mathbf{h}_k$ is the $M \times 1$ vector channel including large and small scale fading between User-k and the BS (α_k and $\mathbf{h}_k$ respectively), x_k is the transmitted data symbol by User-k and $\mathbf{n}_d$ emphasizes the noise on the received *data* signal.

B. Employing an MMSE Receiver at the BS

In this paper the BS employs an MMSE receiver $\mathbf{G}_\kappa \in \mathbb{C}^{1 \times N_r}$ to estimate the data symbol transmitted by User-κ. We recall that the MMSE receiver aims at minimizing the mean-square error between the estimate $\mathbf{G}_\kappa \mathbf{y}$ and the transmitted symbol x_κ:

$$\mathbf{G}_\kappa \triangleq \arg\min_{\mathbf{G}} \mathbb{E}\{\text{MSE}\} = \arg\min_{\mathbf{G}} \mathbb{E}\{|\mathbf{G}\mathbf{y} - x_\kappa|^2\}. \tag{6}$$

When the BS employs a naïve receiver, the estimated channel is taken as if it was the actual channel:

$$\mathbf{G}_\kappa^{\text{naive}} = \alpha_\kappa \sqrt{P_\kappa}\hat{\mathbf{h}}_\kappa^H (\alpha_\kappa^2 P_\kappa \hat{\mathbf{h}}_\kappa \hat{\mathbf{h}}_\kappa^H + \sigma_d^2 \mathbf{I})^{-1}. \tag{7}$$

As we shall see, this receiver does not minimize the MSE.

C. Determining the Actual MMSE Receiver Matrix

In this section we determine the MMSE receiver matrix $\mathbf{G}_k$ that the BS should use to demodulate the received data signal such that the data estimation error for User-κ is minimized taking explicitly account that the BS has access only to the estimated channels $\hat{\mathbf{h}}_\kappa$, as opposed to the naïve receiver that minimizes the MSE only when perfect channel estimation is assumed. To this end, we consider the MSE of the estimated data symbols of the tagged User-κ: obtained from the signal model of (5) using a receiver vector $\mathbf{G}_\kappa$:

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_1, \ldots, \mathbf{h}_K\right) = \mathbb{E}_{x,\mathbf{n}_d}\{|\mathbf{G}_\kappa \mathbf{y} - x_\kappa|^2\} = \\
&= \mathbb{E}_{x,\mathbf{n}_d}\left|(\mathbf{G}_\kappa \alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} - 1)x_\kappa + \sum_{k \neq \kappa}^{K} \mathbf{G}_\kappa \alpha_k \mathbf{h}_k \sqrt{P_k} x_k + \mathbf{G}_\kappa \mathbf{n}_d\right|^2 \\
&= \mathbb{E}_{x,\mathbf{n}_d}\left|(\mathbf{G}_\kappa \alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} - 1)x_\kappa\right|^2 + \sum_{k \neq \kappa}^{K} P_k \mathbb{E}_{x,\mathbf{n}_d}|\mathbf{G}_\kappa \alpha_k \mathbf{h}_k x_k|^2 \\
&\quad + \mathbb{E}_{x,\mathbf{n}_d}\left|\mathbf{G}_\kappa \mathbf{n}_d\right|^2,
\end{aligned} \tag{8}$$

where we utilized that $\mathbb{E}\{x_k\} = 0$ and $\mathbb{E}\{\mathbf{n}_d\} = \mathbf{0}$.

Additionally, utilizing $\mathbb{E}\{x_k x_k^*\} = 1$ and $\mathbb{E}\{\mathbf{n}_d \mathbf{n}_d^H\} = \sigma_d^2 \mathbf{I}_{N_r}$, we have:

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_1, \ldots, \mathbf{h}_K\right) = \\
&\left|\mathbf{G}_\kappa \alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} - 1\right|^2 + \sum_{k \neq \kappa}^{K} P_k |\mathbf{G}_\kappa \alpha_k \mathbf{h}_k|^2 + \sigma_d^2 \mathbf{G}_\kappa \mathbf{G}_\kappa^H,
\end{aligned} \tag{9}$$

from which our first result follows.

Result 1. *When the BS uses the receiver vector $\mathbf{G}_\kappa$, the* MSE *of the received data symbols of the tagged user κ assuming perfect channel state information at the base station is:*

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_\kappa\right) = \mathbb{E}_{\mathbf{h}_1, \ldots, \mathbf{h}_{\kappa-1}, \mathbf{h}_{\kappa+1}, \ldots, \mathbf{h}_K}\{\text{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_1, \ldots, \mathbf{h}_K\right)\} \\
&= \alpha_\kappa^2 P_\kappa \mathbf{G}_\kappa \mathbf{h}_\kappa \mathbf{h}_\kappa^H \mathbf{G}_\kappa^H - \alpha_\kappa \sqrt{P_\kappa}(\mathbf{G}_\kappa \mathbf{h}_\kappa + \mathbf{h}_\kappa^H \mathbf{G}_\kappa^H) + 1 \\
&\quad + \sigma_d^2 \mathbf{G}_\kappa \mathbf{G}_\kappa^H + \sum_{k \neq \kappa}^{K} \alpha_k^2 P_k \mathbf{G}_\kappa \mathbf{C}_k \mathbf{G}_\kappa^H.
\end{aligned} \tag{10}$$

Although this result is useful, we need an expression for the MSE as a function of $\hat{\mathbf{h}}$, rather than $\mathbf{h}$.

Result 2. *The* MSE *of the received data symbols of the tagged user κ as a function of the estimated channel at the BS is:*

$$\begin{aligned}
&\text{MSE}\left(\mathbf{G}_\kappa, \hat{\mathbf{h}}_\kappa\right) = \mathbb{E}_{\mathbf{h}_\kappa | \hat{\mathbf{h}}_\kappa} \text{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_\kappa\right) \qquad (11) \\
&= \alpha_\kappa^2 P_\kappa \mathbf{G}_\kappa (\mathbf{D}_\kappa \hat{\mathbf{h}}_\kappa \hat{\mathbf{h}}_\kappa^H \mathbf{D}_\kappa^H + \mathbf{Q}_\kappa)\mathbf{G}_\kappa^H + \sigma_d^2 \mathbf{G}_\kappa \mathbf{G}_\kappa^H + 1 \\
&\quad + \sum_{k \neq \kappa}^{K} \alpha_k^2 P_k \mathbf{G}_\kappa \mathbf{C}_k \mathbf{G}_\kappa^H - \alpha_\kappa \sqrt{P_\kappa}(\mathbf{G}_\kappa \mathbf{D}_\kappa \hat{\mathbf{h}}_\kappa + \hat{\mathbf{h}}_\kappa^H \mathbf{D}_\kappa^H \mathbf{G}_\kappa^H).
\end{aligned}$$

Using these results, we are in the position of deriving the optimal MU MIMO receiver vector for User-κ:

Proposition 3. *The optimal $\mathbf{G}_\kappa^\star$ can be derived as:*

$$\begin{aligned}
&\mathbf{G}_\kappa^\star = \alpha_\kappa \sqrt{P_\kappa}\hat{\mathbf{h}}_\kappa^H \mathbf{D}_\kappa^H \cdot \qquad (12) \\
&\cdot \left(\alpha_\kappa^2 P_\kappa \left(\mathbf{D}_\kappa \hat{\mathbf{h}}_\kappa \hat{\mathbf{h}}_\kappa^H \mathbf{D}_\kappa^H + \mathbf{Q}_\kappa\right) + \sum_{k \neq \kappa}^{K} \alpha_k^2 P_k \mathbf{C}_k + \sigma_d^2 \mathbf{I}\right)^{-1}.
\end{aligned}$$

IV. Determining the MSE of the Received Data Symbols with Optimal $\mathbf{G}^\star$

In the case of proper antenna spacing, the channel covariance matrices can be modeled as $\mathbf{C}_\kappa = c_\kappa \mathbf{I}$, which for $k = \kappa$ implies $\mathbf{D}_\kappa = d_\kappa \mathbf{I}$, $\mathbf{Q}_\kappa = q_\kappa \mathbf{I}$. In this case, the MSE as a function of the estimated channel can be obtained as follows.

Lemma 4. *In the case of uncorrelated antennas at the* BS, *when the BS employs the optimal receiver $\mathbf{G}_\kappa^\star$, the mean square error of the received data symbols can be expressed as:*

$$\begin{aligned}
&\text{MSE}\left(\hat{\mathbf{h}}_\kappa\right) = -2\alpha_\kappa \sqrt{P_\kappa} g_\kappa d_\kappa ||\hat{\mathbf{h}}_\kappa||^2 + 1 + \\
&g_\kappa^2 \cdot \left(\alpha_\kappa^2 P_\kappa d_\kappa^2 ||\hat{\mathbf{h}}_\kappa||^4 + \left(\alpha_\kappa^2 P_\kappa q_\kappa + \sum_{k \neq \kappa}^{K} \alpha_k^2 P_k c_k + \sigma_d^2\right) ||\hat{\mathbf{h}}_\kappa||^2\right)
\end{aligned}$$

where

$$g_\kappa \triangleq \frac{\alpha_\kappa \sqrt{P_\kappa} d_\kappa}{\alpha_\kappa^2 P_\kappa \left(d_\kappa^2 ||\hat{\mathbf{h}}_\kappa||^2 + q_\kappa\right) + \sum_{k \neq \kappa}^{K} \alpha_k^2 P_k c_k + \sigma_d^2}. \tag{13}$$

We can now derive the unconditional MSE from $\text{MSE} = \mathbb{E}_{\hat{\mathbf{h}}_\kappa} \text{MSE}\left(\hat{\mathbf{h}}_\kappa\right)$ based on the distribution of $\hat{\mathbf{h}}_\kappa$ which we recall from (3) as $\hat{\mathbf{h}}_\kappa \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_\kappa\right)$.

V. Calculating the Unconditional MSE

To calculate the unconditional MSE, notice that the $\text{MSE}(\hat{\mathbf{h}}_\kappa)$ depends on $\hat{\mathbf{h}}_\kappa$ only through $||\hat{\mathbf{h}}_\kappa||^2$. Thus, we can conveniently introduce $Y_\kappa \triangleq ||\hat{\mathbf{h}}_\kappa||^2$, substitute g_κ into (13) and, by inspecting (13), introduce the following notations:

$$T_1 \triangleq g_\kappa^2 \left(\alpha_\kappa^2 P_\kappa d_\kappa^2 ||\hat{\mathbf{h}}_\kappa||^4\right) = \frac{s_\kappa^2 Y_\kappa^2}{(b_\kappa + s_\kappa Y_\kappa)^2},$$

where we introduced the notation $s_\kappa \triangleq d_\kappa^2 p_\kappa$, $p_\kappa \triangleq \alpha_\kappa^2 P_\kappa$, $\sigma_\kappa^2 \triangleq \sum_{k \neq \kappa}^K \alpha_\kappa^2 P_\kappa c_\kappa + \sigma_d^2$ and $Y_\kappa \triangleq ||\hat{\mathbf{h}}_\kappa||^2$ and $b_\kappa \triangleq q_\kappa p_\kappa + \sigma_\kappa^2$. Similarly:

$$T_2 \triangleq g_\kappa^2 \left(\alpha_\kappa^2 P_\kappa q_\kappa + \sum_{k \neq \kappa}^K \alpha_k^2 P_k c_k + \sigma_d^2\right) ||\hat{\mathbf{h}}_\kappa||^2 = \frac{b_\kappa s_\kappa Y_\kappa}{(b_\kappa + s_\kappa Y_\kappa)^2},$$

$$T_3 \triangleq 2 d_\kappa \alpha_\kappa \sqrt{P_\kappa} ||\hat{\mathbf{h}}_\kappa||^2 \cdot g_\kappa = \frac{2 s_\kappa Y_\kappa}{b_\kappa + s_\kappa Y_\kappa}.$$

We can now prove the following proposition, which will serve as the basis for numerical evaluations.

Proposition 5. *The unconditional* MSE *of the received data symbols of User-κ when the BS uses the optimal $\mathbf{G}_\kappa^\star$ receiver is as follows.*

$$\begin{aligned} \text{MSE} = &\frac{b_\kappa \left(e^{\frac{b_\kappa}{s_\kappa r}} (b_\kappa + N_r s_\kappa r) E_{in}\left(N_r, \frac{b_\kappa}{s_\kappa r}\right) - s_\kappa r\right)}{s_\kappa^2 r^2} \\ &+ \frac{N_r \left(e^{\frac{b_\kappa}{s_\kappa r}} \left(b_\kappa + (1+N_r) s_\kappa r\right) E_{in}\left(1+N_r, \frac{b_\kappa}{s_\kappa r}\right) - s_\kappa r\right)}{s_\kappa r} \\ &- 2 \cdot e^{\frac{b_\kappa}{s_\kappa r}} N_r E_{in}\left(1 + N_r, \frac{b_\kappa}{s_\kappa r}\right) + 1, \end{aligned} \tag{14}$$

where $E_{in}(n, z) \triangleq \int_1^\infty e^{-zt}/t^n \; dt$ is a standard exponential integral function.

VI. Numerical Results and Concluding Remarks

Table I: System Parameters

Parameter	Value
Number of antennas	$N_r = 2, 4, 8, 10, 20, 50, 100, 500$
Path Loss of tagged User-κ	$\alpha = 40, 45, 50$ dB
Number of pilot and data symbols	$\tau_p = 1;\; \tau_d = 11$
Power budget	$\tau_p P_p + \tau_d P = P_{tot}$ =250 mW.

In this section we consider a single cell single user MIMO system, in which the mobile terminal is equipped with a single transmit antenna, whereas the BS employs N_r receive antennas. Note that the performance characteristics of the proposed MMSE receiver as compared with the naïve receiver are similar in the multi-user MIMO case from the perspective of the tagged user, since the proposed receiver treats the multi-user interference as noise according to (12). The key input parameters to this system that are necessary to obtain numerical results using the MSE derivation in this paper (ultimately relying on Proposition 5) are listed in Table I.

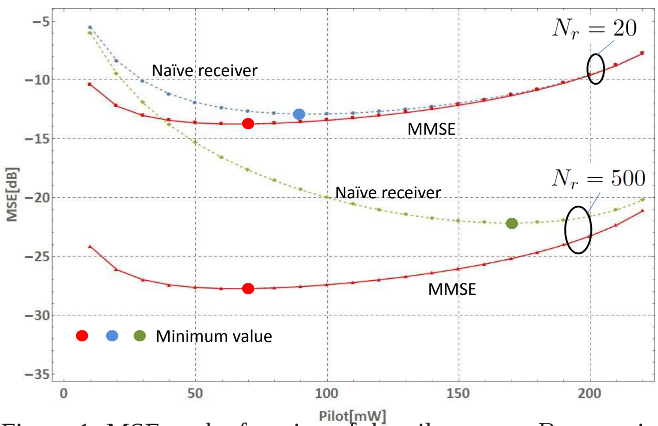

Figure 1: MSE as the function of the pilot power P_p assuming a fixed pilot+data power budget with $N_r = 20$ and $N_r = 500$ number of antennas when using the naïve receiver and the MMSE receiver.

Figure 1 compares the performance of the system in which the number of antennas at the BS grows large ($N_r = 500$). As expected, given a fix sum power budget of $\tau_p P_p + \tau_d P = P_{tot}$ =250 mW, the optimal pilot-data power allocation becomes non trivial as it depends on the number of antennas, path loss and the employed receiver structure. The minimum value of the MSE in all cases are marked with a dot, which clearly indicate that the achievable minimum MSE with this power budget is significantly lower when employing the MMSE receiver.

Figure 2 shows the achievable minimum MSE value and the optimal pilot power setting as the function of the number of antennas at the base station. First, notice that the gain in terms of achievable minimum MSE increases as the number of antennas increases.

For example, at $N_r = 500$ the gain is around 6 dB. Interestingly, the pilot power setting that minimizes the MSE does not depend on the number of antennas when using the MMSE receiver, whereas it increases with the number of antennas in the case of the naïve receiver. The intuitive explanation for this is that in the case of uncorrelated antennas, according to equation (3), the diagonal elements of the covariance of the CSI error does not depend on the number of antennas, although the size of the matrix does. Thus, the pilot-data ration when using the MMSE receiver does not depend on the number of antennas, as opposed to the naïve receiver case, which does not minimize the MSE. The formal proof of this phenomenon is planned for future work.

Appendix I: Proof of Result 1

From (9) it follows, that focusing on the tagged User-κ:

$$\begin{aligned} \text{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_\kappa\right) &= \mathbb{E}_{\mathbf{h}_1, \ldots, \mathbf{h}_{\kappa-1}, \mathbf{h}_{\kappa+1}, \ldots, \mathbf{h}_K} \text{MSE}\left(\mathbf{G}_\kappa, \mathbf{h}_1, \ldots, \mathbf{h}_K\right) \\ &= \left|\mathbf{G}_\kappa \alpha_\kappa \mathbf{h}_\kappa \sqrt{P_\kappa} - 1\right|^2 + \sum_{k, k \neq \kappa} \alpha_k^2 P_k \mathbb{E}_{\mathbf{h}_k} |\mathbf{G}_\kappa \mathbf{h}_k|^2 + \sigma_d^2 \mathbf{G}_\kappa \mathbf{G}_\kappa^H. \end{aligned}$$

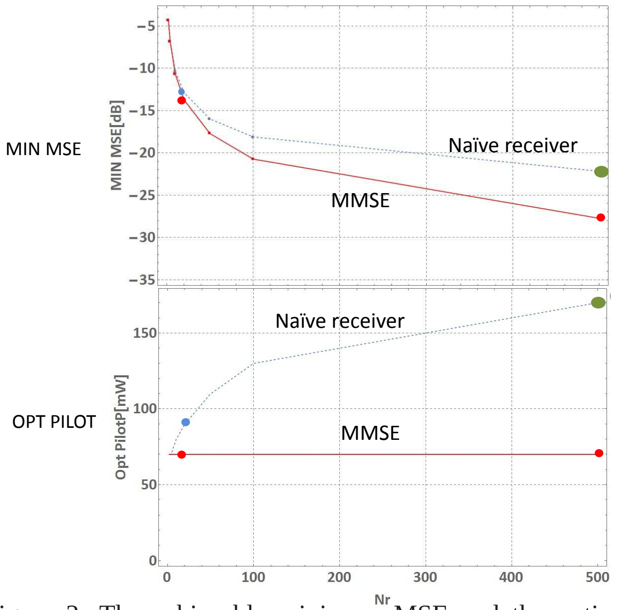

Figure 2: The achievable minimum MSE and the optimum pilot power as the function of the number of the base station antennas when employing the naïve receiver and the MMSE receiver. The dots in the figure correspond to the case of $N_r = 20$ and $N_r = 500$ antennas.

Recognizing that [10]

$$\left|\mathbf{G}\mathbf{h}\alpha\sqrt{P}-1\right|^2 = \alpha^2 P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H - \alpha\sqrt{P}(\mathbf{G}\mathbf{h}+\mathbf{h}^H\mathbf{G}^H)+1,$$

and $\mathbb{E}_{\mathbf{h}_k}|\mathbf{G}_\kappa\mathbf{h}_k|^2 = \mathbf{G}_\kappa\mathbb{E}_{\mathbf{h}_k}|\mathbf{h}_k|^2\mathbf{G}_\kappa^H = \mathbf{G}_\kappa\mathbf{C}_k\mathbf{G}_\kappa^H$, the result follows.

Appendix II: Proof of Result 2

Utilizing $(\mathbf{h}_\kappa \mid \hat{\mathbf{h}}_\kappa) \sim \mathbf{D}_\kappa\hat{\mathbf{h}}_\kappa + \mathcal{CN}\left(\mathbf{0}, \mathbf{Q}_\kappa\right)$, where $\mathbf{D}_\kappa = \mathbf{C}_\kappa{\mathbf{R}_\kappa}^{-1}$, $\mathbf{R}_\kappa = \mathbf{C}_\kappa + \mathbf{C}_\kappa^{\mathbf{w}}$ and $\mathbf{Q}_\kappa = \mathbf{C}_\kappa - \mathbf{C}_\kappa{\mathbf{R}_\kappa}^{-1}\mathbf{C}_\kappa$, and, by averaging over $\mathbf{h}_\kappa|\hat{\mathbf{h}}_\kappa$, and following the technique proposed in [10], the result follows.

Appendix III: Proof of Proposition 3

To derive the optimal $\mathbf{G}_\kappa$, we rewrite MSE $\left(\mathbf{G}_\kappa, \hat{\mathbf{h}}_\kappa\right)$ in quadratic form of $(\mathbf{x}\mathbf{A}\mathbf{x}^H - \mathbf{x}\mathbf{B} - \mathbf{B}^H\mathbf{x}^H + 1)$:

$$\text{MSE}\left(\mathbf{G}_\kappa, \hat{\mathbf{h}}_\kappa\right) = 1 - \alpha_\kappa\sqrt{P_\kappa}\hat{\mathbf{h}}_\kappa^H\mathbf{D}_\kappa^H\mathbf{G}_\kappa^H - \underbrace{\mathbf{G}_\kappa}_{\mathbf{x}}\underbrace{\alpha_\kappa\sqrt{P_\kappa}\mathbf{D}_\kappa\hat{\mathbf{h}}_\kappa}_{\mathbf{B}}$$
$$+\,\mathbf{G}_\kappa\underbrace{\left(\alpha_\kappa^2 P_\kappa\left(\mathbf{D}_\kappa\hat{\mathbf{h}}_\kappa\hat{\mathbf{h}}_\kappa^H\mathbf{D}_\kappa^H + \mathbf{Q}_\kappa\right) + \sum_{k\neq\kappa}^{K}\alpha_k^2 P_k\mathbf{C}_k + \sigma_d^2\mathbf{I}\right)}_{\mathbf{A}}\mathbf{G}_\kappa^H$$

Based on this quadratic form, the optimal receiver $(\mathbf{x}^\star = \mathbf{B}^H\mathbf{A}^{-1})$ is as in (12).

Appendix IV: Proof of Lemma 4

If $\mathbf{C}_\kappa = c_\kappa\mathbf{I}$, implying $\mathbf{D}_\kappa = d_\kappa\mathbf{I}$, $\mathbf{Q}_\kappa = q_\kappa\mathbf{I}$ and the optimal $\mathbf{G}_\kappa^\star$ can be written as:

$$\mathbf{G}_\kappa^\star = \frac{\alpha_\kappa\sqrt{P_\kappa}d_\kappa}{\alpha_\kappa^2 P_\kappa\left(d_\kappa^2||\hat{\mathbf{h}}_\kappa||^2 + q_\kappa\right) + \sum_{k\neq\kappa}^{K}\alpha_k^2 P_k c_k + \sigma_d^2}\,\hat{\mathbf{h}}_\kappa^H$$
$$\triangleq g_\kappa \cdot \hat{\mathbf{h}}_\kappa^H. \tag{15}$$

Substituting $\mathbf{G}_\kappa^\star$ into the MSE of Result 2 gives the lemma.

Appendix V: Proof of Proposition 5

Recognizing that Y_κ is Gamma distributed, the density function of $Y_\kappa \forall\kappa$ is given by (dropping the index κ for convenience):

$$f_Y(x) = \frac{r^{-N_r}x^{N_r-1}e^{-x/r}}{(N_r-1)!} \quad x > 0. \tag{16}$$

Proposition (5) follows from Lemma (4) taking the average of MSE $\left(\hat{\mathbf{h}}_\kappa\right)$ using the the following integrals:

$$\int_{x=0}^{\infty} T_1 f_{Y_\kappa}(x)dx =$$
$$\frac{N_r\left(e^{\frac{b_\kappa}{s_\kappa r}}\left(b_\kappa+(1+N_r)s_\kappa r\right)E_{in}\left(1+N_r, \frac{b_\kappa}{s_\kappa r}\right) - s_\kappa r\right)}{s_\kappa r},$$

$$\int_{x=0}^{\infty} T_2 f_{Y_\kappa}(x)dx = \frac{b_\kappa\left(e^{\frac{b_\kappa}{s_\kappa r}}\left(b_\kappa + N_r s_\kappa r\right)E_{in}\left(N_r, \frac{b_\kappa}{s_\kappa r}\right) - s_\kappa r\right)}{s_\kappa^2 r^2},$$

$$\int_{x=0}^{\infty} T_3 f_{Y_\kappa}(x)dx = 2\cdot e^{\frac{b_\kappa}{s_\kappa r}} N_r E_{in}\left(1+N_r, \frac{b_\kappa}{s_\kappa r}\right),$$

where $E_{in}(n,z) \triangleq \int_1^\infty e^{-zt}/t^n\ dt$ is a standard exponential integral function.

References

[1] N. Jindal and A. Lozano, "A Unified Treatment of Optimum Pilot Overhear in Multipath Fading Channels," *IEEE Transactions on Communications*, vol. 58, no. 10, pp. 2939–2948, October 2010.

[2] N. Sun and J. Wu, "Maximizing spectral efficiency for high mobility systems with imperfect channel state information," *IEEE Transactions on Wireless Communication*, vol. 13, no. 3, pp. 1462–1470, March 2014.

[3] T. Kim and J. G. Andrews, "Optimal Pilot-to-Data Power Ratio for MIMO-OFDM," *IEEE Globecom*, 2005.

[4] T. Marzetta, "How Much Training is Needed for Multiuser MIMO ?" 40^{th} *Asilomar Conference*, October 2006.

[5] K. M. Z. Islam, T. Y. Al-Naffouri, and N. Al-Dhahir, "On Optimum Pilot Design for Comb-Type OFDM Transmission over Doubly-Selective Channels," *IEEE Trans. Comm.*, vol. 59, no. 4, pp. 930–935, April 2011.

[6] G. Fodor, P. D. Marco, and M. Telek, "On the Impact of Antenna Correlation on the Pilot-Data Balance in Multiple Antenna Systems," in *IEEE ICC'15*, June 2015, pp. 1–6.

[7] A. Sezgin, E. A. Jorswieck, and E. Costa, "LDC in MIMO Ricean Channels: Optimal Transmit Strategy with MMSE Detection," *IEEE Trans. on Signal Processing*, vol. 56, no. 1, pp. 313–328, January 2011.

[8] N. Kim, Y. Lee, and H. Park, "Performance Analysis of MIMO System with Linear MMSE Receiver," *IEEE Trans. On Wireless Comm.*, vol. 7, no. 11, pp. 4474–4478, November 2008.

[9] E. Eraslan, B. Daneshrad, and C.-Y. Lou, "Performance Indicator for MIMO MMSE Receivers in the Presence of Channel Estimation Error," *IEEE Wireless Commun. Letters,*, vol. 2, no. 2, pp. 211–214, 2013.

[10] G. Fodor, P. D. Marco, and M. Telek, "Performance Analysis of Block and Comb Type Channel Estimation for Massive MIMO Systems," in *First International Conference on 5G*, Finland, November 2014.